

# Friends-of-Friends Groups and Clusters in the 2SLAQ Catalogue

S. Farrens<sup>1,2\*</sup>, F.B. Abdalla<sup>2</sup>, E.S. Cypriano<sup>3</sup>, C. Sabiu<sup>2</sup> & C. Blake<sup>4</sup>

<sup>1</sup>*Institut de Ciències de l'Espai, IEEC-CSIC, Campus UAB, F. de Ciències, Torre C5 par-2, Barcelona 08193, Spain*

<sup>2</sup>*Department of Physics & Astronomy, University College London, Gower Street, London WC1E 6BT, UK.*

<sup>3</sup>*Departamento de Astronomia, IAG-USP, Rua do Matão 1226, 05508-090, São Paulo, Brazil.*

<sup>4</sup>*Centre for Astrophysics & Supercomputing, Swinburne University of Technology, P.O. Box 218, Hawthorn, VIC 3122, Australia.*

Accepted ???, Received ???; in original form 29 June 2011

## ABSTRACT

We present a catalogue of galaxy groups and clusters selected using a friends-of-friends algorithm with a dynamic linking length from the 2dF-SDSS and QSO (2SLAQ) luminous red galaxy survey. The linking parameters for the code are chosen through an analysis of simulated 2SLAQ haloes. The resulting catalogue includes 313 clusters containing 1,152 galaxies. The galaxy groups and clusters have an average velocity dispersion of  $\bar{\sigma}_v = 467.97 \text{ km s}^{-1}$  and an average size of  $\bar{R}_{clt} = 0.78 \text{ Mpc } h^{-1}$ . Galaxies from regions of one square degree and centred on the galaxy clusters were downloaded from the Sloan Digital Sky Survey Data Release 6 (SDSS DR6). Investigating the photometric redshifts and cluster red-sequence of these galaxies shows that the galaxy clusters detected with the FoF algorithm are reliable out to  $z \sim 0.6$ . We estimate masses for the clusters using their velocity dispersions. These mass estimates are shown to be consistent with 2SLAQ mock halo masses. Further analysis of the simulation haloes shows that clipping out low richness groups with large radii improves the purity of catalogue from 52% to 88%, while retaining a completeness of 94%. Finally, we test the two-point correlation function of our cluster catalogue. We find a best-fitting power law model,  $\xi(r) = (r/r_0)^\gamma$ , with parameters  $r_0 = 24 \pm 4 \text{ Mpc } h^{-1}$  and  $\gamma = -2.1 \pm 0.2$ , which are in agreement with other low redshift cluster samples and consistent with a  $\Lambda$ CDM universe.

**Key words:** Galaxies: Clusters: General - Galaxies: Distances and Redshifts

## 1 INTRODUCTION

Galaxy clusters are the largest bound objects in the Universe and are important structures for examining the distribution of matter and how this evolves with time. Investigating how the mass-function of galaxy clusters changes with time provides an effective method for constraining cosmological parameters (Press & Schechter 1974; Peebles 1993; Weller et al. 2002; Weller & Battye 2003; Haiman et al. 2001).

The first galaxy cluster was unknowingly detected by Charles Messier, who recorded the positions of 11 nebulae in the Virgo cluster in the 18<sup>th</sup> century. Later evidence for the existence of clusters of galaxies was provided in the work of Harlow Shapley and Adelaide Ames in the 1930s (Shapley & Ames 1932). They severely undermined the idea that galaxies are randomly distributed throughout the Universe. Probably the most significant figure in the pioneering

of galaxy cluster detection was George Abell (Abell 1958). Abell surveyed around three quarters of the sky using photographic plates, which meant that he had to identify the locations of galaxy overdensities by eye. To avoid including field galaxies, Abell defined galaxy clusters as regions 1.5 Mpc  $h^{-1}$  in radius (the Abell Radius) containing fifty or more galaxies that are no more than two magnitudes fainter than the third brightest member of the group. Remarkably, later studies have confirmed a large number of Abell's clusters as being genuine bound structures. Due to the simplicity of this two-dimensional approach, however, serious problems can arise from *e.g.* inhomogeneities, photometric errors and projection effects. Bahcall & Soneira (1983) found an excess of power in the angular correlation function using Abell clusters and van Haarlem et al. (1997) show, using N-body simulations, that  $\sim 1/3$  of Abell's clusters have incorrect richness estimates and  $\sim 1/3$  of Abell richness class  $R \geq 1$  clusters are missed. Around thirty years after Abell's work hybrid photo-digital surveys were able to improve upon the photographic plate method for detecting clusters of galaxies

\* Email: farrens@ieec.uab.es

(Maddox et al. 1990). It was, however, the advent of digital CCD surveys that brought about significant advances, a good example being surveys such as the SDSS (York et al. 2000), 2dFGRS (Colless 1999), 6dF (Jones et al. 2006) and the ongoing GAMA (Driver et al. 2008). Large sky surveys like SDSS signify a major step forward in obtaining galaxy data, however analysing that data can be approached and interpreted in many different ways. Automated algorithms supply a means of reducing subjectivity in the analysis of large data sets. Finally, the availability of increasingly detailed simulations in recent years has enabled more powerful tests of the completeness and reliability of cluster catalogues.

In the last few decades many different algorithms have been developed to find galaxy clusters. The *Counts in Cells* method (Couch et al. 1991; Lidman & Peterson 1996) looks for enhancements of galaxy surface density in a given area. *Percolation* methods group together galaxies that are separated by a distance on the sky less than a given threshold distance (Efsthathiou et al. 1988; Davis et al. 1985; Dalton et al. 1997; Ramella et al. 2002). This technique was originally applied to redshift surveys by Huchra & Geller (1982). An extended version of Huchra and Geller’s friend-of-friends algorithm was later developed to deal with photometric redshifts by Botzler et al. (2004). The changes implemented in the extended friend-of-friends algorithm were necessary to account for the large errors associated with photometric redshifts (Betzler et al. 2004). *Matched Filter* techniques model the spatial and luminosity distributions of galaxies in a cluster to generate a cluster likelihood map (Postman et al. 1996; Kim et al. 2002). *Voronoi Tessellation* decomposes the galaxies in a region of space into discrete points surrounded by cells facilitating the identification of overdensities (Kim et al. 2002; Lopes et al. 2004; Ramella et al. 2001). Another method is to look for a *Red Sequence* in the colour-magnitude relation of galaxy clusters (Gladders & Yee 2000; Koester et al. 2007). The *Cut and Enhance* method makes colour and colour-colour cuts to produce subsamples of galaxies in different redshift ranges (Goto et al. 2002). A review of various methods and techniques used for the optical detection of galaxy clusters is provided by Gal (2006).

The Percolation or friend-of-friends (FoF) algorithm is the cluster finding method of interest for the purposes of this paper. The 2dF-SDSS and QSO (2SLAQ) Luminous Red Galaxy Survey, which is subset of photometrically selected luminous red galaxies (LRGs) from SDSS (Cannon et al. 2006), is a good starting place for the use of the FoF algorithm because of its relatively small size and the abundance of both spectroscopic and photometric data available.

This paper is the first in a series of two papers exploring the detection and analysis of galaxy clusters in the 2SLAQ catalogue and will focus on utilising the spectroscopic data available in the catalogue. The second paper, Farrens et al. (in prep), will investigate detecting clusters using photometric redshifts estimated from the SDSS photometry in the 2SLAQ catalogue and compare the results with those provided in this paper. This comparison will demonstrate the reliability of the FoF technique to detect clusters with photometric redshifts on a consistent data set. Large data sets are required if one is to do cosmology with galaxy clusters. The fastest way of obtaining this data is through photometric surveys. Upcoming surveys like the Dark Energy Survey

(DES), Euclid and Planck will obtain such photometric data for large numbers of galaxies across the sky and since it is not feasible to obtain spectra for all of these objects within the observing time frames, it is therefore necessary to develop reliable methods for detecting clusters using photometric data. Photometric cluster catalogues made from surveys such as these can cover a larger volume than spectroscopic catalogues and can thus be used to probe the shape of the mass function at higher masses.

The following section provides some background on the 2SLAQ survey. Section 3 describes the Huchra and Geller Friends-of-friends method in detail and how we implemented the algorithm. Section 4 provides a description of the mock catalogue used to calibrate the linking parameters. Section 5 presents the resulting groups and clusters found using the FoF algorithm and the analysis made on these results. Finally, section 6 shows the 2-point correlation function of the clusters. We present a sample of the cluster catalogue in the appendix of this paper.

## 2 THE 2DF-SDSS AND QSO (2SLAQ) LUMINOUS RED GALAXY SURVEY

2SLAQ is a spectroscopic survey of around 15000 potential luminous red galaxies (LRGs) in the redshift range  $0.45 \leq z \leq 0.7$  (Cannon et al. 2006). The target LRGs were selected photometrically from the Sloan Digital Sky Survey (York et al. 2000). 2SLAQ also includes a lower resolution survey of faint quasi-stellar objects (QSOs) or quasars. Observations began in March 2003 using the Two-degree Field instrument (2dF) on the 3.9m Anglo-Australian Telescope.

The 2SLAQ criteria for selecting galaxies from the Sloan Digital Sky Survey consisted in placing limits on the magnitudes, colours and star-galaxy separations of the SDSS data. The magnitude limits were imposed by requiring that:

$$17.5 \leq i_{deV} - A_i < 19.8 \quad (1)$$

$$i_{fibre} < 21.2 \quad (2)$$

where  $i_{deV}$  is the total magnitude based on a fit of each galaxy to a de Vaucouleurs profile,  $A_i$  is the extinction in the  $i$ -band and  $i_{fibre}$  is the flux contained within the aperture of a spectroscopic fibre in the  $i$ -band. These limits enabled the selection of bright LRGs out to  $z \sim 0.8$  and eliminated objects too diffuse to produce useful spectra. The colour cuts were such that:

$$0.5 < g - r < 3.0 \quad (3)$$

$$r - i < 2.0 \quad (4)$$

$$c_{\parallel} \equiv 0.7(g - r) + 1.2(r - i - 0.18) > 1.6 \quad (5)$$

$$d_{\perp} \equiv (r - i) - \frac{(g - r)}{8.0} > 0.5 \quad (6)$$

where  $c_{\parallel}$  eliminates later-type galaxies and  $d_{\perp}$  selects early-type galaxies at increasingly high redshift. Equations 3 and 4 ensure that only objects near the main locus of LRGs are selected. The star-galaxy separation criteria was:

$$i_{psf} - i_{model} > 0.2 + 0.2(20.0 - i_{deV}) \quad (7)$$

$$\text{radius}_{deV(i)} > 0.2'' \quad (8)$$

where  $i_{psf}$  is the  $i$ -band magnitude of an isolated star fitted with a point spread function (PSF) model,  $i_{model}$  is the model  $i$ -band magnitude, which is a good proxy for the PSF, and  $\text{radius}_{deV}$  is the de Vaucouleurs radius, which is the effective radius in the model magnitude. These limits eliminate the majority of stellar contamination.

### 3 FRIENDS-OF-FRIENDS METHOD

#### 3.1 The Huchra and Geller Friends-of-Friends Algorithm

The Huchra and Geller friends-of-friends algorithm was developed to search for groups of galaxies in the magnitude limited CFA1 redshift survey (Huchra & Geller 1982; Huchra et al. 1983; Geller & Huchra 1983). This technique is commutative and utilises only a galaxy's right ascension, declination and redshift to detect structure by finding galaxies that are separated by a distance less than some threshold,  $D_L$ , and that have a velocity difference less than some threshold,  $V_L$ . The threshold values are chosen depending on the properties of the galaxies in the catalogue. For simplicity  $D_L$  and  $V_L$  can be set as fixed values, however this may lead to some selection effects being ignored. Huchra & Geller (1982) adopt a method that compensates for the variation in the completeness of the galaxy luminosity function as a function of redshift.

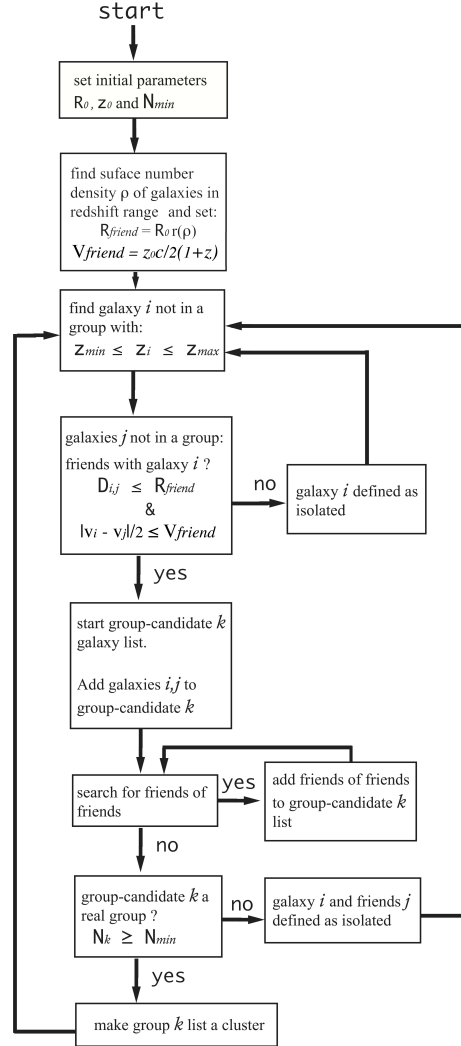
#### 3.2 Choice of Dynamic Friends-of-Friends Linking

The friends-of-friends algorithm used in this paper follows the Huchra & Geller (1982) method in most respects. The only difference lies in the linking parameters, which change with the surface number density of galaxies at a given redshift. This dynamic linking length is more appropriate to this sample of galaxies as it does not assume a magnitude limit pre-selection, which would remove large numbers of galaxies usable for cluster detection and prevent any clusters at the highest redshifts from being found. The choice of this type of linking parameter has the drawback of allowing for a more complex selection function that will render interpretation less straight forward. A fluxogram of the Dynamic Friends-of-Friends (hereafter DFoF) algorithm is shown in fig.1.

Following the steps shown in fig.1: First the initial values of linking length,  $R_0$ , redshift linking,  $z_0$ , and the minimum number of galaxy members needed to form a group/cluster,  $N_{min}$ , are chosen. To compensate for selection effects,  $R_0$  is varied via equation 9:

$$R_{friend}(z) \propto R_0 \left( \frac{dN}{dz} \frac{dz}{dV} \frac{1}{A_{sky}} \right)^{-\frac{1}{2}} \quad (9)$$

where  $z$  is redshift of each galaxy for which the variable linking length is calculated,  $dN(z)/dz$  is the surface number density of galaxies in the redshift range covered by the galaxy catalogue ( $z_{min} \leq z \leq z_{max}$ ),  $dV/dz$  is the differential comoving volume and  $A_{sky}$  is the fraction of sky area covered in the catalogue relative to the total sky area. Note that the limits  $z_{min}$  and  $z_{max}$  can be modified if one does not wish to use the full redshift range of the catalogue.  $z_0$  is



**Figure 1.** Fluxogram of the FoF algorithm with dynamic linking length. Based on Huchra & Geller (1982).

converted into a velocity linking parameter,  $v_{friend}(z)$ , via equation 10:

$$v_{friend}(z) = \frac{z_0 c}{2(1+z)} \quad (10)$$

where  $c$  is the speed of light. The factor of 2 in the denominator accounts for the fact that we want to be able to link galaxies that have velocities approaching along the line of sight to those that have velocities receding along the line of sight.  $v_{friend}(z)$  will therefore be representative of the velocity dispersions of the clusters being examined.

A galaxy  $i$  is selected from the catalogue that has not yet been assigned to a group and lies in the redshift range  $z_{min} \leq z \leq z_{max}$ . The projected distance between galaxy  $i$  and second galaxy  $j$ ,  $D_{ij}$ , is calculated using:

$$D_{ij} = \cos^{-1}(\sin(\delta_i)\sin(\delta_j) + \cos(\delta_i)\cos(\delta_j)\cos(\alpha_i - \alpha_j)) \quad (11)$$

where  $\alpha$  and  $\delta$  are right ascension and declination respectively.

The two galaxies are linked together (*i.e.* are friends) if

they satisfy the conditions in equations 12 and 13:

$$D_{ij} \leq R_{friend}(z) \quad (12)$$

$$\frac{|v_i - v_j|}{2} \leq v_{friend}(z) \quad (13)$$

A group-candidate  $k$  is formed that includes galaxy  $i$  and its friends. A search is then made around the galaxies linked to  $i$ . This process is repeated until no further friends are found. The group-candidate  $k$  is defined as a real group if it satisfies equation 14:

$$N_k \geq N_{min} \quad (14)$$

For the purposes of this paper  $N_{min} \geq 3$ . This is a reasonable assumption when dealing with LRGs, which are not common objects.

#### 4 MOCK CATALOGUE

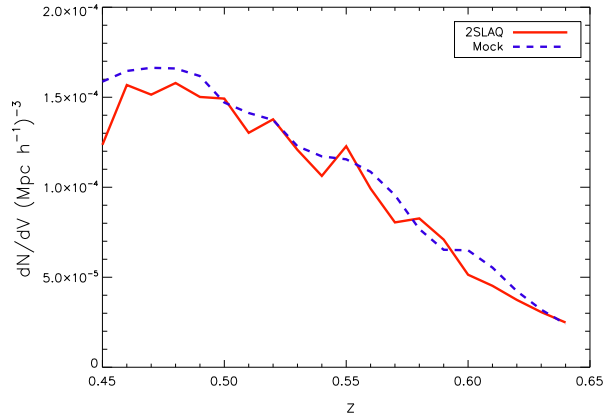
To determine the optimum values of  $R_0$  and  $z_0$  for the DFoF code, a mock galaxy catalogue that simulates the 2SLAQ catalogue was produced. This mock catalogue contains a distribution of 7,651,076 dark matter haloes and 824,704 galaxies across an octant of the sky.

The halo catalogue was derived from the Horizon 4 $\pi$  simulation (Teyssier et al. 2009; Prunet et al. 2008). This is a  $\Lambda$ CDM dark matter  $N$ -body simulation using WMAP 3 cosmology with a  $2h^{-1}$  Gpc periodic box on a grid of 40,963 cells. The  $7 \times 10^{10}$  particles were evolved using the Particle Mesh scheme of the RAMSES code (Teyssier 2002) on an adaptively refined grid (AMR) with around  $1.4 \times 10^{11}$  cells, reaching a formal resolution of 262,144 cells in each direction ( $\sim 7h^{-1}$  kpc comoving). The simulation covers a sufficiently large volume to compute a full-sky dark matter distribution, while resolving Milky-Way size haloes with more than 100 particles and exploring small scales deeply into the non-linear regime. The dark matter distribution in the simulation was integrated in a light cone out to redshift 1, around an observer located at the centre of the simulation box. The underlying cosmology for WMAP 3 is:  $\Omega_M = 0.24$ ,  $\Omega_\Lambda = 0.76$ ,  $\Omega_b = 0.042$ ,  $n = 0.958$ ,  $H_0 = 73$  and  $\sigma_8 = 0.77$ .

Gravitationally bound haloes of dark matter are selected using the spherical overdensity HOP<sup>1</sup> method of Eisenstein & Hut (1998). HOP is based on a hybrid approach in which the local density field is first obtained by smoothing the density field with an SPH-like kernel using the  $n$  nearest neighbours. Then the particles above a given threshold are linked with their highest density neighbours until, after several ‘‘hops’’, they are connected to the one particle with the highest density within the region above the threshold. All particles linked to the local density maximum are identified as a group.

The haloes were then populated using the Halo Occupation Distribution (HOD), where the number of galaxies residing within each halo is drawn from a probability,  $P(N|M)$ , that a dark matter halo of mass  $M$  will host  $N$  galaxies.

The first moment of  $P(N|M)$  is the mean number of



**Figure 2.** Number density as a function of redshift for the mock halo catalogue (blue dashed line) and the 2SLAQ catalogue (red solid line).

**Table 1.** T

he HOD parameters derived from the MegaZ-LRG sample using the methodology of Blake et al. (2008). The small differences between these parameters and those of Blake et al. (2008) reflect the difference in the assumed underlying cosmology.

Redshift slice	$\sigma_{cut}$	$\log\left(\frac{M_0}{M_\odot/h}\right)$	$\beta$	$\log\left(\frac{M_{cut}}{M_\odot/h}\right)$
$0.45 < z < 0.50$	0.618	13.88	1.41	12.96
$0.50 < z < 0.55$	0.469	13.99	1.54	13.00
$0.55 < z < 0.60$	0.554	14.16	1.66	13.23
$0.60 < z < 0.65$	0.675	14.43	1.56	13.60

galaxies as a function of halo mass and it is usually parameterised as a sum of a central and a satellite components (Kraivtsov et al. 2004; Tinker 2007).

The probability that a halo contains a central galaxy is given by:

$$\langle N_c | M \rangle = 0.5 \left[ 1 + \operatorname{erf} \left( \frac{\log_{10}(M/M_{cut})}{\sigma_{cut}} \right) \right] \quad (15)$$

and the number of satellite galaxies is obtained from a Poisson sampling of:

$$\langle N_s | M \rangle = \left( \frac{M}{M_0} \right)^\beta \quad (16)$$

The HOD model used is that of Blake et al. (2008) and includes several derived parameters that were computed at various redshifts bins between  $0.4 < z < 0.7$ . Table 1 lists the HOD parameters in each redshift bin. For the purposes of this paper an evolving HOD model is constructed that smoothly interpolates between the four Blake et al. (2008) redshift bins as shown in fig.2. This figure shows that our smoothly evolving HOD model, based on Blake et al. (2008), succeeds in reproducing the observed number density of the 2SLAQ data.

The radial positions of satellite galaxies within a halo are assigned according to the NFW profile (Navarro et al. 1996). Specifically, it is assumed that the mass inside a given radius, properly normalised, represents the probability of

<sup>1</sup> <http://cmb.as.arizona.edu/~eisenste/hop/hop.html>



containing a galaxy. Thus, integrating to the halo boundary would yield a probability of one and therefore all galaxies would be placed within this radius. The angular position relative to the halo centre is chosen randomly for each galaxy.

In our model, all dark matter particles are assumed to be in approximately spherical virialised haloes. The velocity of a dark matter particle is the sum of two terms,

$$v = v_{\text{vir}} + v_{\text{halo}}, \quad (17)$$

the first is due to the velocity of the particle about the centre of mass of its parent halo, and the second is due to the motion of the centre of mass of the parent.

Consider the first term,  $v_{\text{vir}}$ . We will assume that virialised haloes are isothermal spheres, so that the distribution of velocities within them is Maxwellian. This is in reasonable agreement with measurements of virial velocities within haloes in numerical simulations. If  $\sigma_{\text{vir}}$  denotes the rms speeds of particles within a halo, then the virial theorem requires that:

$$\frac{Gm}{r} \propto \sigma_{\text{vir}}^2 \propto \frac{H(z)^2}{2} \Delta_{\text{vir}}^{1/3}(z) \left( \frac{3m}{4\pi\rho_{\text{crit}}(z)} \right)^{2/3} \quad (18)$$

where the final proportionality comes from the fact that all haloes have the same density whatever their mass:  $m/r^3 \propto \Delta_{\text{vir}} \rho_{\text{crit}}$ . This shows that  $\sigma_{\text{vir}} \propto m^{1/3}$ : the more massive haloes are expected to be ‘hotter’. At fixed mass, the constant of proportionality depends on time and cosmology, and on the exact shape of the density profile of the halo. A convenient fitting formula is provided by Bryan & Norman (1998):

$$\sigma_{\text{vir}}(m, z) = 102.5 g_{\sigma} \Delta_{\text{vir}}^{1/6}(z) \left( \frac{H(z)}{H_0} \right)^{1/3} \left( \frac{m}{10^{13} M_{\odot}/h} \right)^{1/3} \quad (19)$$

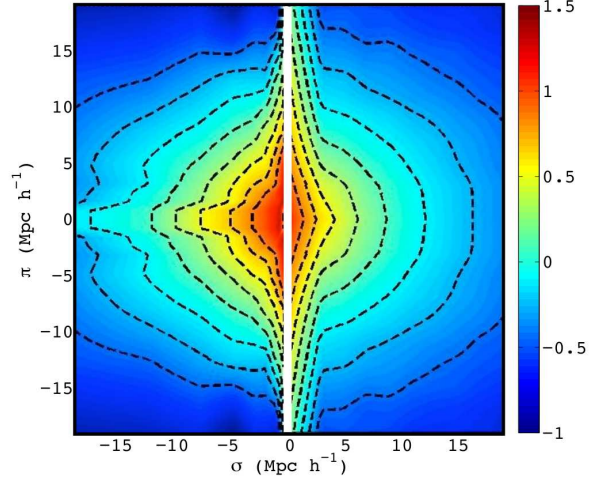
where  $g_{\sigma} = 0.9$ , and

$$\Delta_{\text{vir}} = 18\pi^2 + 82x - 39x^2, \quad \text{with } x = \Omega(z) - 1 \quad (20)$$

and  $\Omega(z) = [\Omega_m (1+z)^3] [H_0/H(z)]^2$ . It has been shown that  $\sigma(M)$  is independent of local environment (Sheth & Diaferio 2001), however it may depend on position within a halo. This will mainly be due to the fact that haloes have complicated density and velocity profiles.

The two-dimensional redshift-space correlation function for the mock catalogue,  $\xi(\sigma, \pi)$ , was calculated with the pair separation decomposed in terms of perpendicular,  $\sigma$ , and parallel,  $\pi$ , distances. Each component of separation was calculated for 20 bins equally spaced in comoving distance between  $0.01 < \sigma, \pi [\text{Mpc} h^{-1}] < 40$ , creating a 400 element grid of measurements. The smooth contours are created by linear interpolation over the grid. Fig.3 shows the calculated 2-D correlation functions for the 2SLAQ catalogue of Ross et al. (2007) and the mock galaxy catalogue. It shows the expected linear squashing at large radii and redshift distortions with ‘finger of God’ elongations at small scales. The colour map represents  $\log(\xi)$ .

Fig.2 and 3 show that the simulation reproduces the properties of 2SLAQ well enough for us to believe that it is sufficient to aid us in calibrating our cluster finding parameters and therefore any limits imposed upon the values of  $R_0$  and  $z_0$  from the simulation will be applicable to the real data. The discrepancy between the real and mock 2SLAQ data in fig.3 at low values of  $\pi$  is due to shot noise (*i.e.* there is a small number of pairs at this scale). There are, however,



**Figure 3.** The two-dimensional redshift-space correlation function for the 2SLAQ catalogue (left half) and the mock catalogue (right half) plotted as a function of transverse,  $\sigma$ , and radial,  $\pi$ , pair separation. The colour map represents  $\log(\xi)$ . The 2SLAQ correlation function is that of Ross et al. (2007).

some caveats that should be taken into account that may cause the simulation to not be fully representative of reality. For example, the effects of things such as the way in which galaxies are added to dark matter haloes and cluster mergers at high redshifts could have influence on the results. In the following section, we discuss how the simulations can bias the results together with the analysis of the results.

## 5 DATA ANALYSIS

### 5.1 Linking Parameter Optimisation

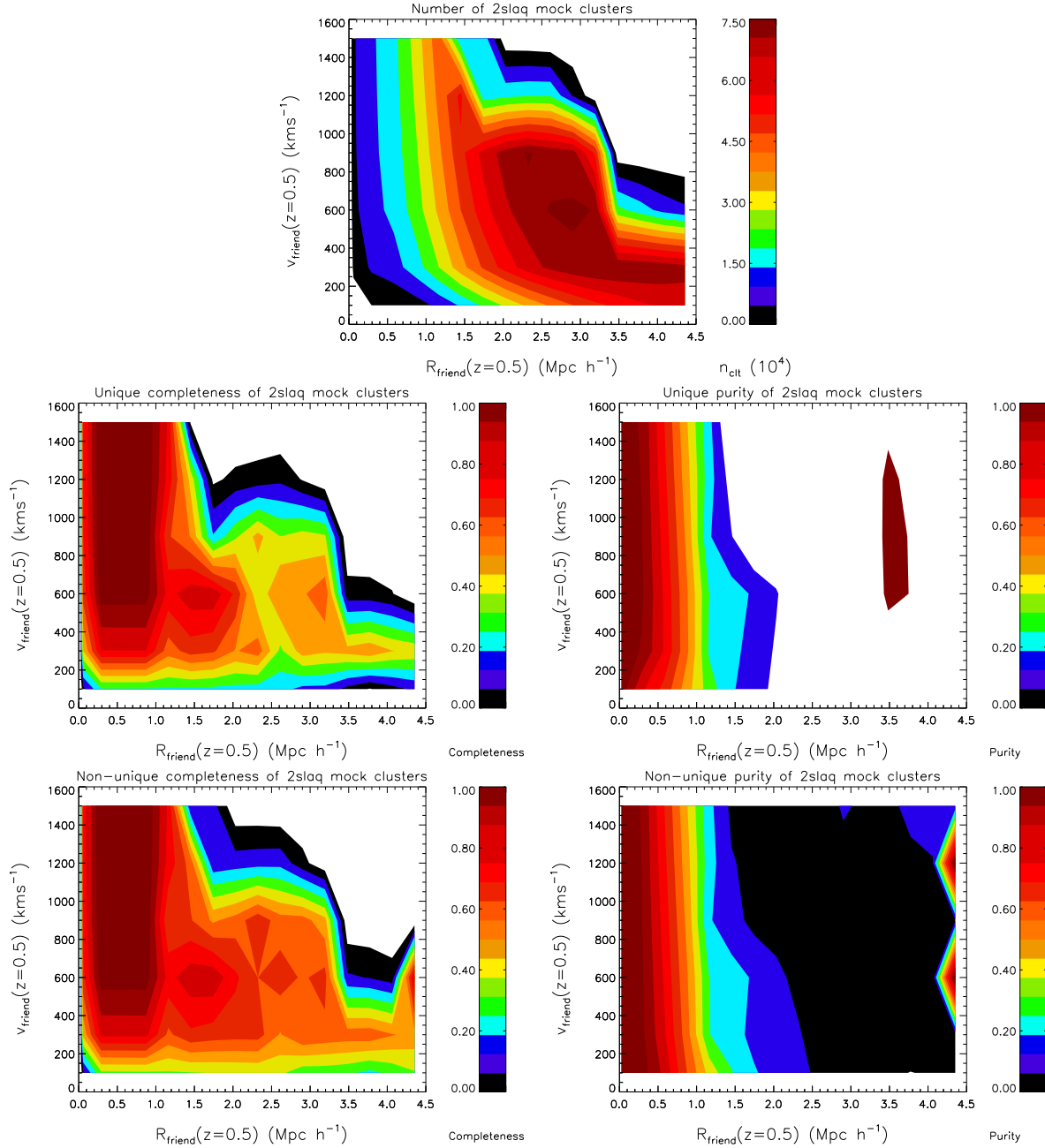
The DFofF code was run on the 2SLAQ simulation with different values of  $R_0$  and  $z_0$  to find the optimum combination of the two. To gain a more physical interpretation of these parameters, hereafter  $R_0$  and  $z_0$  are expressed in terms of  $R_{\text{friend}}(z)$  and  $v_{\text{friend}}(z)$  at  $z = 0.5$  using equations 9 and 10. The parameters were varied in the ranges:

$$0 \leq R_{\text{friend}}(z = 0.5) \leq 4.2 \text{ (Mpc } h^{-1}) \quad (21)$$

$$100 \leq v_{\text{friend}}(z = 0.5) \leq 1500 \text{ (kms}^{-1}) \quad (22)$$

The choice of parameters is made based on the completeness, purity and total number of clusters in the resulting catalogues.

In order to determine the completeness and purity of the DFofF clusters relative to the simulation haloes, a membership matching code was implemented. The technique involves looking at the galaxy members assigned to each cluster by the DFofF code and matching these to the original halo member galaxies. The clusters are examined in descending order of richness to ensure that the largest clusters are the first to be matched to mock haloes. DFofF clusters may contain contributions from several mock haloes, therefore each cluster is matched to the mock halo with the highest number of shared members. Thus, if two haloes are merged into one cluster only one of the two haloes will be matched.



**Figure 4.** Number of clusters detected in the mock halo catalogue as a function of  $R_{\text{friend}}(z=0.5)$  and  $v_{\text{friend}}(z=0.5)$  (top panel). Unique completeness (middle left panel) and purity (middle right panel) of the DFOF clusters relative to mock haloes as a function of  $R_{\text{friend}}(z=0.5)$  and  $v_{\text{friend}}(z=0.5)$ . Non-unique completeness (bottom left panel) and purity (bottom right panel) of the DFOF clusters relative to mock haloes as a function of  $R_{\text{friend}}(z=0.5)$  and  $v_{\text{friend}}(z=0.5)$ .

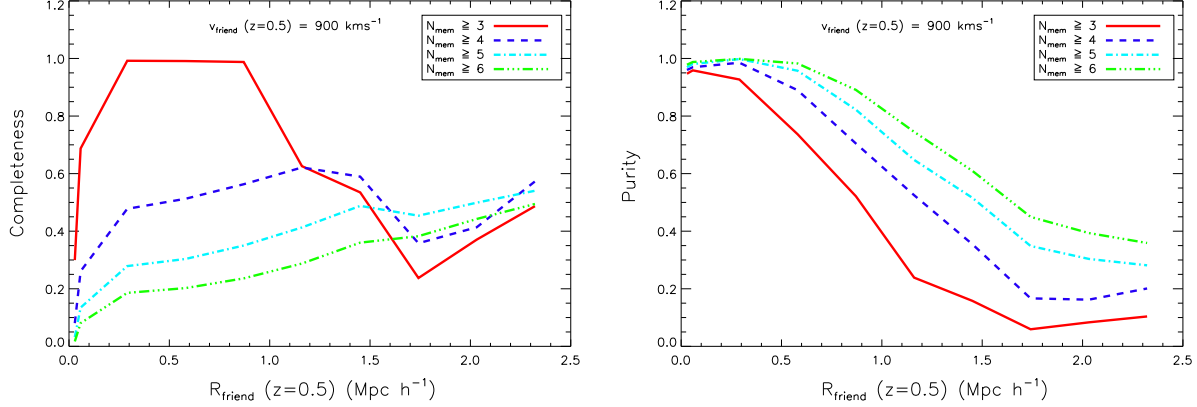
For the purposes of this paper we examine two matching scenarios: a) a strict unique-matching regime, b) a less strict multiple-matching regime. In the first case each cluster is uniquely matched to one halo and any cluster that corresponds to a halo that has already been matched, which will be of equal or lesser richness, will be ignored. In the latter case multiple clusters are allowed to match to the same halo, which increases the completeness. The completeness and purity are defined via equations 23 and 24.

$$\text{Completeness} = \frac{N_{\text{matches}}}{N_{\text{haloes}}} \quad (23)$$

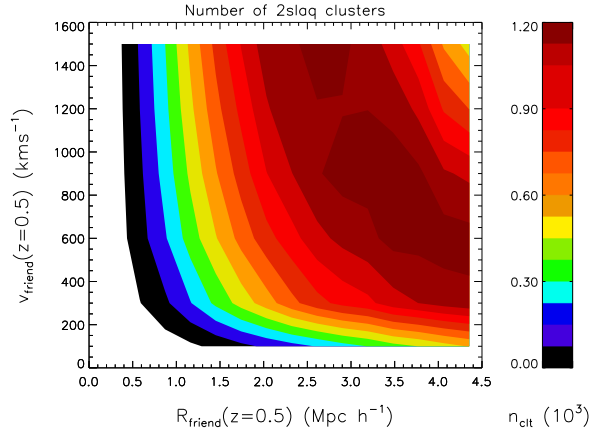
$$\text{Purity} = \frac{N_{\text{matches}}}{N_{\text{clusters}}} \quad (24)$$

Where  $N_{\text{matches}}$  is the total number of unique or multiple matches,  $N_{\text{clusters}}$  is the total number of DFOF clusters found and  $N_{\text{haloes}}$  is the total number of mock haloes.  $N_{\text{haloes}}$  is calculated ignoring all mock haloes with less than 3 galaxy members as this is the detection limit of the DFOF code.

Fig.4 shows the total number of clusters found in the 2SLAQ mock (top panel), unique completeness (middle left panel), unique purity (middle right panel), non-unique com-



**Figure 5.** Completeness (left panel) and purity (right panel) as a function of  $R_{\text{friend}}(z = 0.5)$  for a fixed value of  $v_{\text{friend}}(z = 0.5) = 900 \text{ km s}^{-1}$ . The red solid line shows clusters with  $N_{\text{mem}} \geq 3$ , the blue dashed line shows clusters with  $N_{\text{mem}} \geq 4$ , the light blue dot-dashed line shows clusters  $N_{\text{mem}} \geq 5$  and the green triple dot-dashed line shows clusters with  $N_{\text{mem}} \geq 6$ .



**Figure 6.** Number of clusters detected in the 2SLAQ catalogue as a function of  $R_{\text{friend}}(z = 0.5)$  and  $v_{\text{friend}}(z = 0.5)$ .

completeness (bottom left panel) and non-unique purity (bottom right panel) as a function of  $R_{\text{friend}}(z = 0.5)$  and  $v_{\text{friend}}(z = 0.5)$ . This figure clearly shows that the DFoF results are primarily dependent on the choice of  $R_{\text{friend}}(z = 0.5)$ . Looking at the overall contour shape in the top panel, one can see that the number of clusters found peaks around  $R_{\text{friend}}(z = 0.5) = 2.8 \text{ Mpc h}^{-1}$  and  $v_{\text{friend}}(z = 0.5) = 900 \text{ km s}^{-1}$  after which clusters are merged together creating highly unphysical structures. This sets an upper limit on both of the parameters, although we would intuitively expect the value of  $R_{\text{friend}}(z = 0.5)$  to be much lower. For  $v_{\text{friend}}(z = 0.5) > 400 \text{ km s}^{-1}$ , the catalogue is fully complete in the range  $0.35 < R_{\text{friend}}(z = 0.5) < 0.87 \text{ Mpc h}^{-1}$  and fully pure out to  $R_{\text{friend}}(z = 0.5) = 0.28 \text{ Mpc h}^{-1}$  for both the unique and non-unique regimes.

Since  $R_{\text{friend}}(z = 0.5)$  has a greater effect on the completeness and purity, we can fix  $v_{\text{friend}}(z = 0.5)$  to the limit imposed by the top panel of fig.4 and examine how the completeness and purity vary with just  $R_{\text{friend}}(z = 0.5)$ . Fig.5 shows the variations in completeness and purity as a function of  $R_{\text{friend}}(z = 0.5)$  for different richness cuts using a fixed value of  $v_{\text{friend}}(z = 0.5) = 900 \text{ km s}^{-1}$ . The richness cuts show what the completeness and purity would look like

if we remove groups that have fewer members than some given threshold. It should be noted that the value of  $N_{\text{haloes}}$  is unchanged and therefore haloes which have fewer members than the richness cut threshold are no longer matched. This figure indicates that the richer clusters are more pure, but less complete, as one would expect. Because it is not possible to produce a catalogue that is 100% complete and 100% pure, it is necessary to choose one or the other, or some compromise between the two requirements. If we choose a catalogue that is  $\sim 100\%$  complete, fig.5 shows it can be cut by richness to improve the purity.

Finally, the DFoF code was run on the real 2SLAQ catalogue to investigate the number of clusters found for a given set of linking parameters. Fig.6 shows the total number of clusters found in the 2SLAQ catalogue as a function of  $R_{\text{friend}}(z = 0.5)$  and  $v_{\text{friend}}(z = 0.5)$ . Both this figure and the top panel of fig.4 have the largest  $N_{\text{clt}}$  contours roughly in the range  $2.5 \leq R_{\text{friend}}(z = 0.5) \leq 4.0 \text{ Mpc h}^{-1}$ .

Therefore in order to obtain a fully complete catalogue with high purity and the largest number of clusters possible, based on the results in fig.4, 5 and 6, values of  $R_{\text{friend}}(z = 0.5) = 0.87 \text{ Mpc h}^{-1}$  and  $v_{\text{friend}}(z = 0.5) = 900 \text{ km s}^{-1}$  were chosen to find groups and clusters in the 2SLAQ catalogue.

These values correspond to a 2SLAQ mock cluster catalogue that is 98% complete and 52% pure.

## 5.2 Basic Results

Running the DFoF algorithm using a linking length of  $R_{\text{friend}}(z = 0.5) = 0.87 \text{ Mpc } h^{-1}$  and a velocity linking parameter of  $v_{\text{friend}}(z = 0.5) = 900 \text{ kms}^{-1}$  on the 13,133 galaxies in the 2SLAQ sample produced a total of 313 groups and clusters containing 1,152 member galaxies. Fig.7 shows the angular position of the centres of these groups and clusters and their distribution with respect to redshift. The cluster centres and redshifts are taken as the average values of each of the galaxy members.

Fig.8 shows the distribution of cluster velocity dispersion,  $\sigma_v$ , as a function of redshift. Fig.9 shows the distribution of cluster size,  $R_{\text{clt}}$ , as a function of redshift. Where  $R_{\text{clt}}$  is defined as the projected distance from the cluster centre to the farthest galaxy member. The ‘x’s indicate clusters with three members, the blue circles indicate clusters with between four and six members and the red squares indicate clusters with seven or more members. These plots illustrate that on average the richest clusters have the largest sizes and velocity dispersions as expected.

## 5.3 Background Subtraction

The DFoF groups and clusters were examined by looking for peaks in the photometric redshift distribution and colour-magnitude relation of SDSS galaxies around the cluster centres. Using an SQL query, regions of one square degree and centred on the FoF cluster candidates were downloaded from the SDSS DR6 catalogue. These squares were downloaded for each of the 313 DFoF groups and clusters and include all SDSS galaxies with photometric redshifts. The properties downloaded for each galaxy were: id, ra, dec,  $u$ -magnitude,  $g$ -magnitude,  $r$ -magnitude,  $i$ -magnitude,  $z$ -magnitude, model  $i$ -magnitude, photometric redshift, photometric redshift error and star likelihood. In each of these squares all galaxies within a  $1 \text{ Mpc } h^{-1}$  radius of the centre were taken as cluster members with background, regardless of redshift, and all galaxies between 3 and 7  $\text{Mpc } h^{-1}$  were used as field galaxies. Galaxies between 1 and 3  $\text{Mpc } h^{-1}$  were ignored. These limits were chosen because, on average, we would not expect galaxies farther than 3  $\text{Mpc } h^{-1}$  from a real cluster centre to be genuine members. Using the DFoF cluster redshifts, the data from the corresponding SDSS squares were stacked in redshift slices of  $\Delta z = 0.05$  in the range  $0.4 \leq z \leq 0.7$ . The total number of groups and clusters with three or more and four or more members in each redshift slice are listed in Table-2.

In each cluster redshift bin the stacked SDSS galaxies were binned by photometric redshift. The number of SDSS galaxies assigned to the field was then subtracted from the number of SDSS galaxies assigned as cluster members plus background in each redshift bin. This subtraction was done taking into account the difference in area between the two regions. Fig.10 shows the number of SDSS background subtracted galaxies as a function of photometric redshift for all groups and clusters. The blue dashed line are the background galaxies, the green dot-dashed line are the foreground galaxies and the red solid line are the background

**Table 2.** Redshift Slices

z Range	$N_{\text{clt}}$ with $N_{\text{mem}} \geq 3$	$N_{\text{clt}}$ with $N_{\text{mem}} \geq 4$
$0.40 \leq z \leq 0.45$	26	10
$0.45 \leq z \leq 0.50$	99	46
$0.50 \leq z \leq 0.55$	85	30
$0.55 \leq z \leq 0.60$	62	24
$0.60 \leq z \leq 0.65$	26	14
$0.65 \leq z \leq 0.70$	8	2

subtracted galaxies. In these plots the background subtraction should average out to zero if there is no overdensity in the field. A sharp peak can clearly be seen in each of the plots at the cluster redshift range. This trend is reliable up to  $z \sim 0.6$  after which the peaks cannot clearly be distinguished. This is a strong indication that the DFoF cluster candidates are genuine structures. The discrepancies seen after  $z \sim 0.6$  are due to the relatively low number of clusters at these redshifts plus a result of poor photometric redshift estimates for more distant objects in the SDSS catalogue.

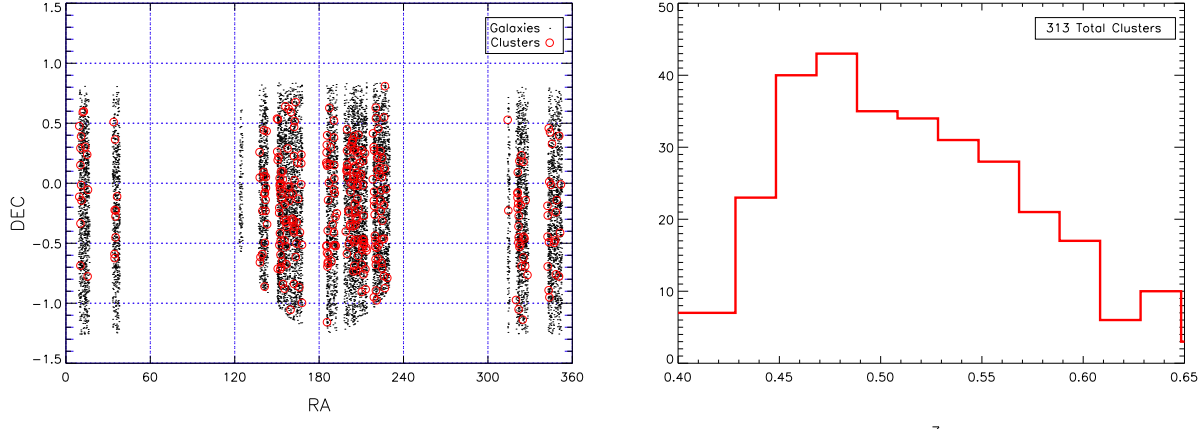
To look for the colour-magnitude relation of the groups and clusters,  $i$ -magnitude vs  $(g - i)$ -colour maps were made for both field galaxies and cluster+field galaxies in each SDSS square. The C-M maps were then stacked by cluster redshift in redshift slices as seen in Table-2. Finally, the stacked field galaxy maps were subtracted from the stacked cluster galaxy maps taking into account the relative areas. Fig.11 shows the background subtracted C-M diagrams for SDSS galaxies. The maps show a clear trend in colour-magnitude space that resembles a cluster red sequence. The cluster red sequence is an observational property whereby cluster galaxies are more red than field galaxies at the same redshift. As with the plots in fig.10 this trend is visible out to  $z \sim 0.6$  and similarly indicates that the structures are genuine.

## 5.4 Mass estimates

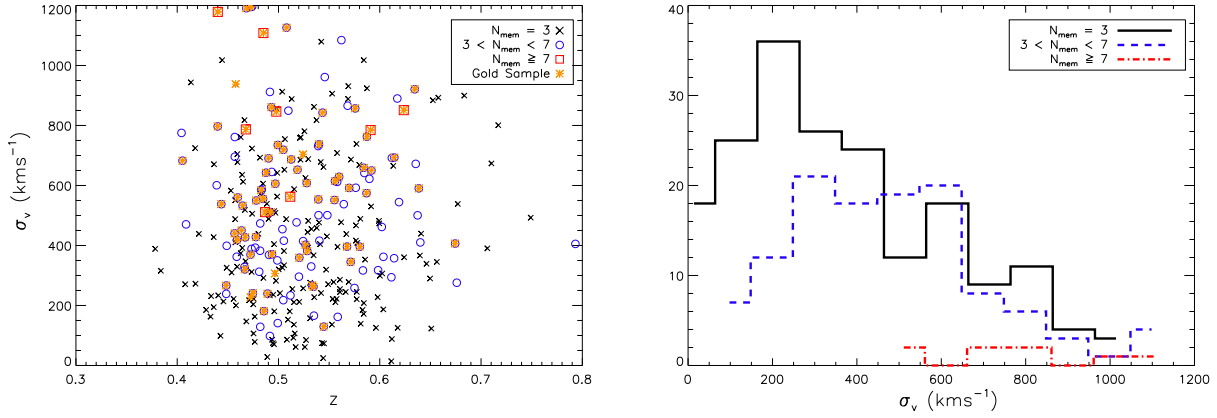
In order to calculate masses for our sample of groups and clusters, we compare mass estimates calculated using equation 19 with the true masses of the 2SLAQ simulation haloes. Masses were calculated for the DFoF catalogue produced from the 2SLAQ mock with linking parameters  $R_{\text{friend}}(z = 0.5) = 0.87 \text{ Mpc } h^{-1}$  and  $v_{\text{friend}}(z = 0.5) = 900 \text{ kms}^{-1}$  using the cluster velocity dispersions (hereafter  $M_{\text{est}}$ ). The membership matching code described in section 5.1 was then implemented to assign to each cluster a mock halo mass (hereafter  $M_{\text{true}}$ ).

Fig.12 shows the average true cluster mass,  $\overline{M}_{\text{true}}$ , as a function of cluster richness,  $N_{\text{mem}}$ , and the average estimated mass,  $\overline{M}_{\text{est}}$ , as a function of true mass. The error bars show the standard error of the mean. The left panel of this plot shows the expected trend between cluster richness and mass. The right panel compares the matched halo masses to those calculated from the cluster velocity dispersions according to equation 19. The deviations from the  $x = y$  trend are principally the result of contaminating galaxies in the richness estimates and the small number of high mass clusters detected. At the low mass range clusters are composed of only 3 or 4 LRGs, hence their velocity dispersions are unreliable and one object that is either missing or interloping in

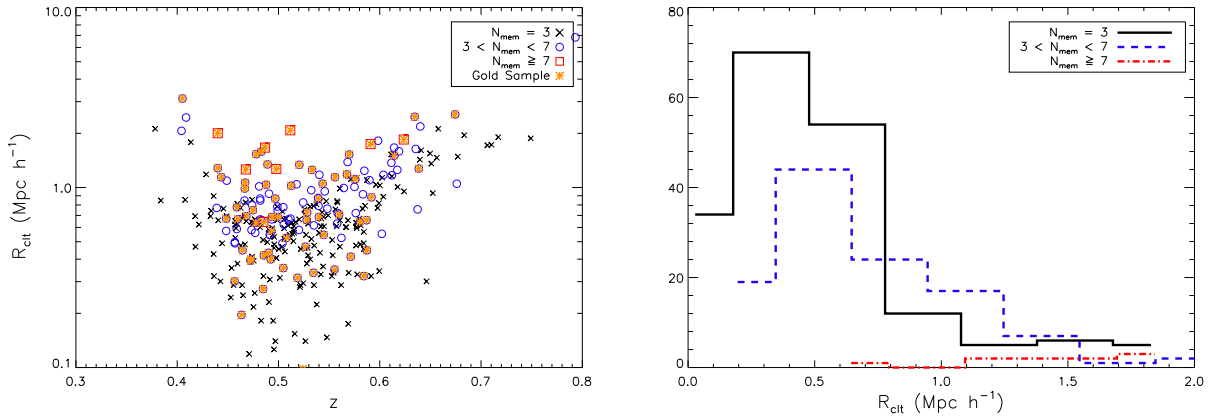




**Figure 7.** Distribution of clusters in RA and Dec (left panel). Histogram of clusters as function of redshift (right panel).



**Figure 8.** Distribution of cluster velocity dispersion,  $\sigma_v$ , as a function of redshift (left panel). The 'x's indicate clusters with less than four members, the blue circles indicate clusters with between four and seven members and the red squares indicate clusters with more than seven members. The gold asterisks highlight the groups and clusters that form part of the gold sample (see §5.5). Histogram of galaxy members as a function of cluster velocity dispersion for the 3 choices of richness as in the left hand plot (right panel).



**Figure 9.** Distribution of cluster size,  $R_{\text{clt}}$ , as a function of redshift (left panel). The 'x's indicate clusters with less than four members, the blue circles indicate clusters with between four and seven members and the red squares indicate clusters with more than seven members. The gold asterisks highlight the groups and clusters that form part of the gold sample (see §5.5). Histogram of galaxy members as a function of cluster size for the 3 choices of richness as in the left hand plot (right panel).



**Figure 10.** Number of SDSS background subtracted galaxies as a function of photometric redshift for all groups and clusters. The blue dashed line are the background galaxies, the green dot-dashed line are the foreground galaxies and the red solid line are the background subtracted galaxies.

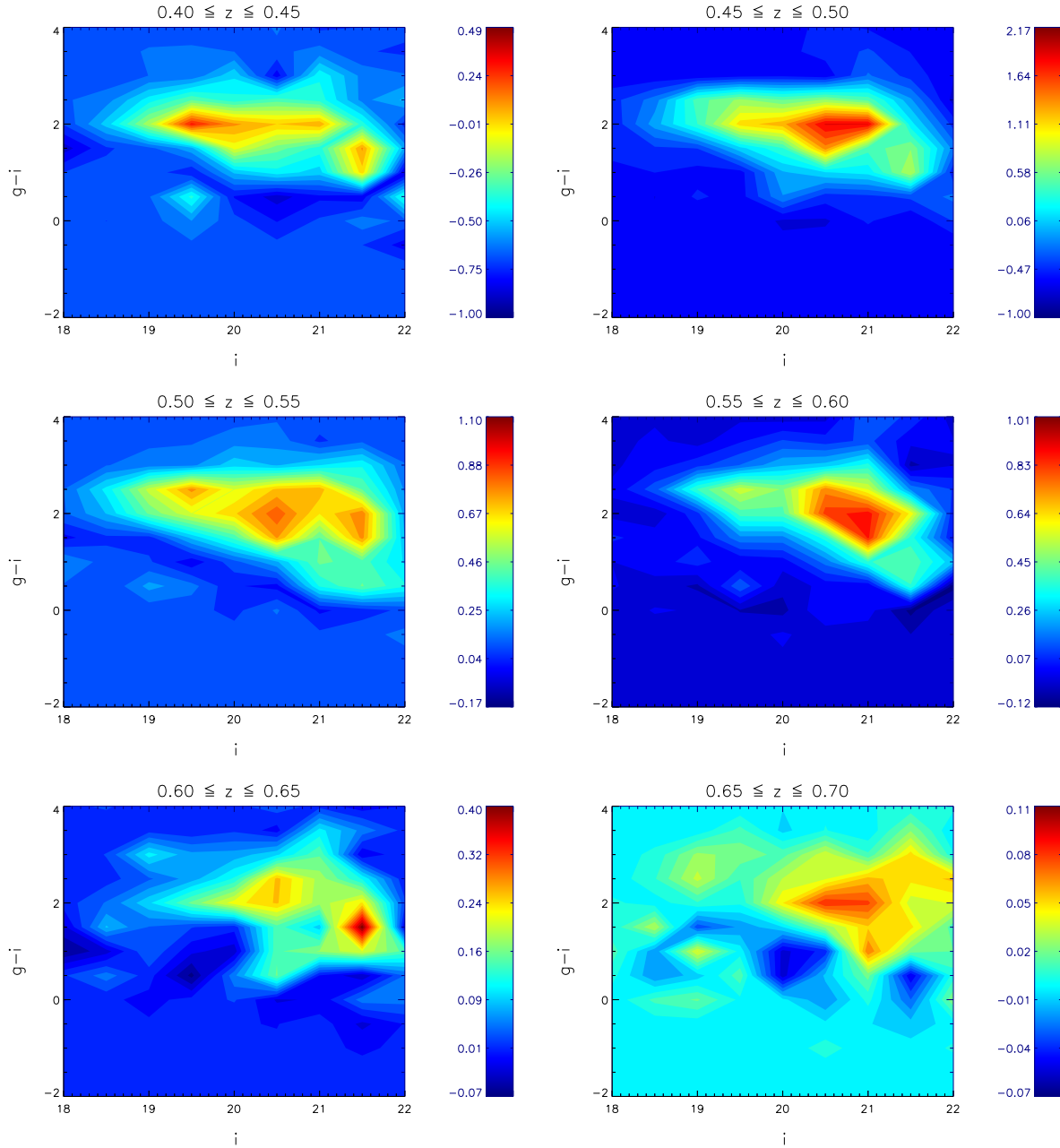
the cluster will have a large impact on the cluster velocity dispersion and hence the mass estimate. At the high mass range cluster velocity dispersions should be more reliable, but there are very few objects that contribute to the average mass estimate. In addition to this, in the high mass range, when all the galaxies in an individual mock halo have been detected perfectly by the DFoF code, the cluster mass from the velocity dispersion tends to be overestimated by around  $10^{0.4} M_{\odot}$ . This is, however, comparable with the size of the error bars in the high mass range of fig.12.

The range of cluster masses is a good match to known masses of massive clusters. Therefore, from this analysis we

can assume that mass estimates made with the real 2SLAQ cluster velocity dispersions will be approximately representative of their true physical masses with a large amount of uncertainty, particularly in the low mass range. This uncertainty could be reduced with a larger sample of clusters and by including all cluster galaxies, rather than just the LRGs, to calculate the velocity dispersions.

## 5.5 Clipping

In section 5.1 we chose to produce a catalogue that optimised the completeness, while maintaining the highest pos-

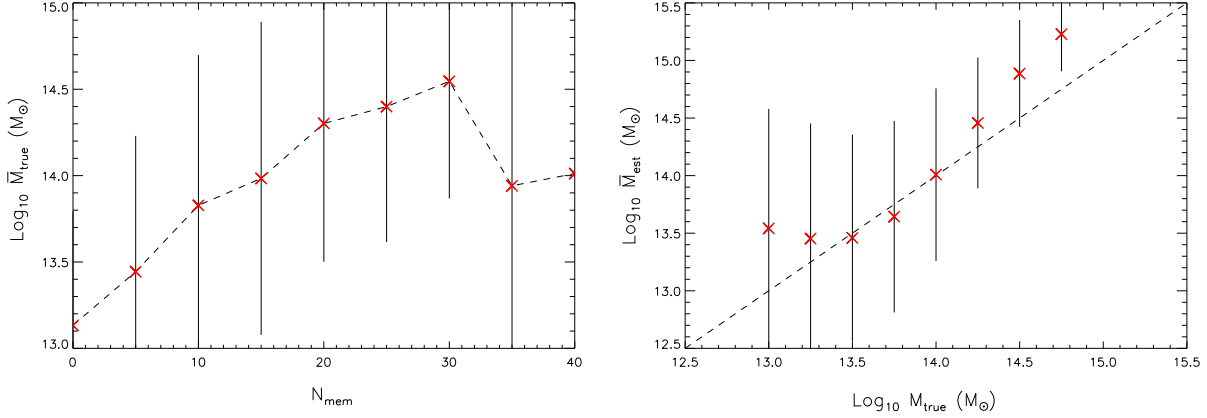


**Figure 11.** Colour-Magnitude diagrams for background subtracted SDSS galaxies. The maps show a clear trend in colour-magnitude space that resembles a cluster red sequence.

sible purity. This choice resulted in a catalogue that was 98% complete and 52 % pure. Fig. 5 shows that clipping groups and clusters with the lowest richnesses will improve the purity, but will reduce the completeness. This plot also shows that the purity is much higher for lower values of  $R_{friend}(z = 0.5)$ , which would correspond to structures that have smaller radial sizes. Therefore, it appears that the purity is mainly affected by the size and richness of the groups and clusters.

In an attempt to improve the purity while preserving the completeness, several new catalogues were made by clipping out groups and clusters with sizes larger than some threshold for a given richness from the original 2SLAQ mock

cluster catalogue. Table 3 shows the total number of groups and clusters with 3 and 4 members for various radial cuts.  $R_{clip}$  is the size threshold in  $\text{Mpc } h^{-1}$  above which all groups and clusters will be removed,  $N_{tot}$  is the total number of groups and clusters for a given  $R_{clip}$ ,  $N_{true}$  is the number of groups and clusters that are matched to 2SLAQ mock haloes for a given  $R_{clip}$  and the ratio  $N_{true}/N_{tot}$  gives a measure for the purity for a given  $R_{clip}$ . As can be seen in the table, when no cuts are made groups of 3 and 4 members are only 38% and 57% pure respectively. Above richness of 4 all groups and clusters are above 80% pure. Therefore, the majority of the contamination in the cluster catalogue arises from these small groups.



**Figure 12.** Average true cluster mass,  $\overline{M}_{true}$ , as a function of cluster richness,  $N_{mem}$  (left panel). Average estimated mass,  $\overline{M}_{est}$ , as a function of true mass (right panel). Error bars are the standard error of the mean.

Taking only groups and clusters with sizes less than  $0.11 \text{ Mpc } h^{-1}$  for  $N_{mem} = 3$  and less than  $0.49 \text{ Mpc } h^{-1}$  for  $N_{mem} = 4$  produces a new catalogue which is 94% complete and 88% pure. These values were chosen because they provide the best improvement to the purity with minimal change to the completeness. Thus, by clipping out clusters with a low richness (*i.e.* few members), but relatively large sizes, we can improve the purity by around 36%, while reducing the completeness by only 4%.

Applying this same clipping procedure to the ‘real’ 2SLAQ groups and clusters, we can separate the catalogue into ‘gold’ and ‘silver’ samples. Where the gold clusters are those that pass the clipping procedure and therefore are the most likely to be genuine. The silver clusters fail the clipping procedure and may still be genuine, but the probability is lower. Out of the 313 total 2SLAQ groups and clusters, 70 are gold and the remaining 243 are silver. The gold sample groups and clusters have an average velocity dispersion of  $\overline{\sigma}_v = 585.28 \text{ kms}^{-1}$  and an average size of  $\overline{R}_{clt} = 0.90 \text{ Mpc } h^{-1}$ , while the silver sample groups and clusters have an average velocity dispersion of  $\overline{\sigma}_v = 430.47 \text{ kms}^{-1}$  and an average size of  $\overline{R}_{clt} = 0.53 \text{ Mpc } h^{-1}$ . The average properties of the gold sample are larger than those of the full (gold+silver) sample as a large fraction of the small structures have been removed.

It should be noted that, although we are confident that our mock is a good representation of the 2SLAQ galaxies, it is difficult to interpret how well this clipping procedure will translate to the real 2SLAQ catalogue. Therefore we do not remove the silver sample clusters, rather we assign to them a lower likelihood of being genuine than the gold sample clusters. The final 2SLAQ cluster catalogue contains an additional column that identifies each cluster as belonging to the gold or silver samples with ‘G’ or ‘S’ respectively. In fig.8 and 9 the groups and clusters that form part of the gold sample are highlighted with gold asterisks.

## 5.6 Cluster Images

Fig. 13 shows SDSS DR7 optical  $g$ ,  $r$  and  $i$ -band colour images of the regions around the DFoF detected clusters CL\_008, CL\_204, CL\_122, CL\_024 and CL\_038. The yellow

**Table 3.** Radial Clipping:  $R_{clip}$  is the size threshold in  $\text{Mpc } h^{-1}$  above which all groups and clusters will be removed,  $N_{tot}$  is the total number of groups and clusters for a given  $R_{clip}$ ,  $N_{true}$  is the number of groups and clusters that are matched to 2SLAQ mock haloes for a given  $R_{clip}$  and the ratio  $N_{true}/N_{tot}$  gives a measure for the purity for a given  $R_{clip}$ .

$R_{clip} \text{ (Mpc } h^{-1})$	$N_{tot}$	$N_{true}$	$N_{true}/N_{tot}$
$N_{mem} = 3$			
-	17382	6741	0.388
0.60	13998	6739	0.481
0.49	11359	6739	0.593
0.39	9367	6739	0.719
0.28	7998	6739	0.842
0.21	7399	6738	0.911
0.14	6993	6708	0.959
<b>0.11</b>	<b>6723</b>	<b>6563</b>	<b>0.976</b>
0.07	5789	5727	0.989
$N_{mem} = 4$			
-	5927	3381	0.571
0.70	4561	3306	0.725
0.60	3987	3131	0.785
<b>0.49</b>	<b>3362</b>	<b>2862</b>	<b>0.851</b>
0.39	2966	2697	0.909
0.32	2728	2582	0.946
0.25	2577	2505	0.972
0.18	2433	2415	0.992
0.11	2116	2108	0.996

circles highlight the positions of the cluster member galaxies, which are labeled with their individual redshifts. These clusters were chosen to sample several different redshift bins as listed in Table 2.

These images show that the galaxies are distributed in a small area on the sky and are close in redshift space. An important point to notice is that the distribution of cluster LRGs in each image varies from spherical to elongated filamentary structures. This is an advantage of the FoF method, which makes no prior assumptions about the shape of the groups and clusters. CL\_204 (top right panel) in particular shows a rather filamentary distribution of cluster LRGs. The



full cluster will be comprised of many galaxy most of which will not be LRGs and therefore the true shape could well be more spherical. Other techniques that have prerequisites on the shape of clusters would not be able to detect this particular object using only the LRGs.

## 6 CORRELATION FUNCTION

The spatial distribution of galaxies and clusters of galaxies contains a wealth of information regarding the underlying cosmological model. The most widely used method in the literature for condensing this information is to measure the autocorrelation function of the positional data. For our purposes we measure the correlation function simply to compare the statistical spatial distribution of our cluster sample to other derived cluster samples.

We calculate the two-point correlation function for the 313 2SLAQ groups and clusters obtained previously. Using the 2SLAQ spectroscopic galaxy catalogue we create a random catalogue, which replicates the angular completeness on the sky, this process is handled straightforwardly in the Healpix<sup>2</sup> package of software. The radial distribution of the random catalogue is obtained from a smooth spline fit to the galaxy redshift distribution,  $n(z)$ . The smoothing ensures the exclusion of large scale structure *e.g.* voids and filaments.

The correlation function is calculated with the Landy and Szalay estimator (Landy & Szalay 1993):

$$\xi = \frac{DD - 2DR + RR}{RR} \quad (25)$$

where  $DD$ ,  $DR$  and  $RR$  are the number of pairs of points in the data,  $D$ , and random,  $R$ , catalogues. The number of pairs are calculated in 14 bins, equally separated in log space, from  $r = 5 - 90$  Mpc  $h^{-1}$ .

The correlation function is usually represented as a power law,  $\xi(r) = (r/r_0)^\gamma$ , thus making a comparison with other works quite straight forward. In the left panel of fig.14 we plot the two-point correlation function of the 2SLAQ clusters (blue squares with error bars) and a best-fit power-law slope (red solid line). The error bars are estimated using the jackknife method, which involves dividing the survey into  $N$  sections with equal area or volume. The variance and mean are estimated from  $N$  measures of our statistic. Each measurement is performed on the survey with region  $i$  removed, where  $i = 1, \dots, N$ .

The jackknife estimate of the variance is (Lupton 1993):

$$\sigma_\xi^2(r_i) = \frac{N_{jk} - 1}{N_{jk}} \sum_{j=1}^{N_{jk}} [\xi_j(r_i) - \bar{\xi}(r_i)]^2 \quad (26)$$

where  $N_{jk}$  is the number of Jackknife samples used and  $r_i$  represents a single bin in our statistic,  $\xi$ . In this analysis we set  $N_{jk} = 20$  sample.

The power law model is best fit with the parameters  $r_0 = 24 \pm 4$  Mpc  $h^{-1}$  and  $\gamma = -2.1 \pm 0.2$ , and the value of the reduced chi-squared is  $\chi_{red}^2 = 0.94$ .

The right panel in fig.14 shows the two-point correlation function of the mock haloes (black squares with error

bars) and a best-fit power-law slope (red solid line). The correlation of the mock haloes is best fit with parameters  $r_0 = 23.05 \pm 0.72$  Mpc  $h^{-1}$  and  $\gamma = -1.95 \pm 0.05$ , which are perfectly consistent with the parameters found for the 2SLAQ groups and clusters. The error bars are measured using the jackknife method as before.

Although the errors are much larger for the real 2SLAQ data, this analysis indicates that the catalogue of groups and clusters found within 2SLAQ using the DFoF code shows the correct level of clustering with respect to the mock halo catalogue. In addition, significant amounts of impurity and/or incompleteness in the cluster catalogue would suppress the correlation length as impurity adds randomly placed structure and incompleteness removes correlated structure. Thus, as both the real and mock 2SLAQ data have consistent correlation lengths, this implies that the amount of incompleteness and impurity in the 2SLAQ cluster catalogue is not significant.

The correlation length of the 2SLAQ LRG sample was calculated by Ross et al. (2007),  $7.45 \pm 0.35$  Mpc  $h^{-1}$  and by Sawangwit et al. (2009),  $7.5 \pm 0.04$  Mpc  $h^{-1}$ , however this value is expected to be larger for clusters of galaxies. Comparing the  $r_0$  value obtained with those found using other low redshift cluster samples shows good agreement, *e.g.*  $r_0 = 26.0 \pm 4.5$  Mpc  $h^{-1}$  (Borgani et al. 1999),  $19.4 \leq r_0 \leq 23.3$  Mpc  $h^{-1}$  (Miller et al. 1999),  $18.8 \pm 0.9$  Mpc  $h^{-1}$  (Collins et al. 2000). See Nichol (2001) for a good review of  $r_0$  values for various cluster samples. The clustering length found here is also consistent with that expected for  $\Lambda$ CDM,  $22 \leq r_0 \leq 27$  Mpc  $h^{-1}$  (Colberg et al. 2000).

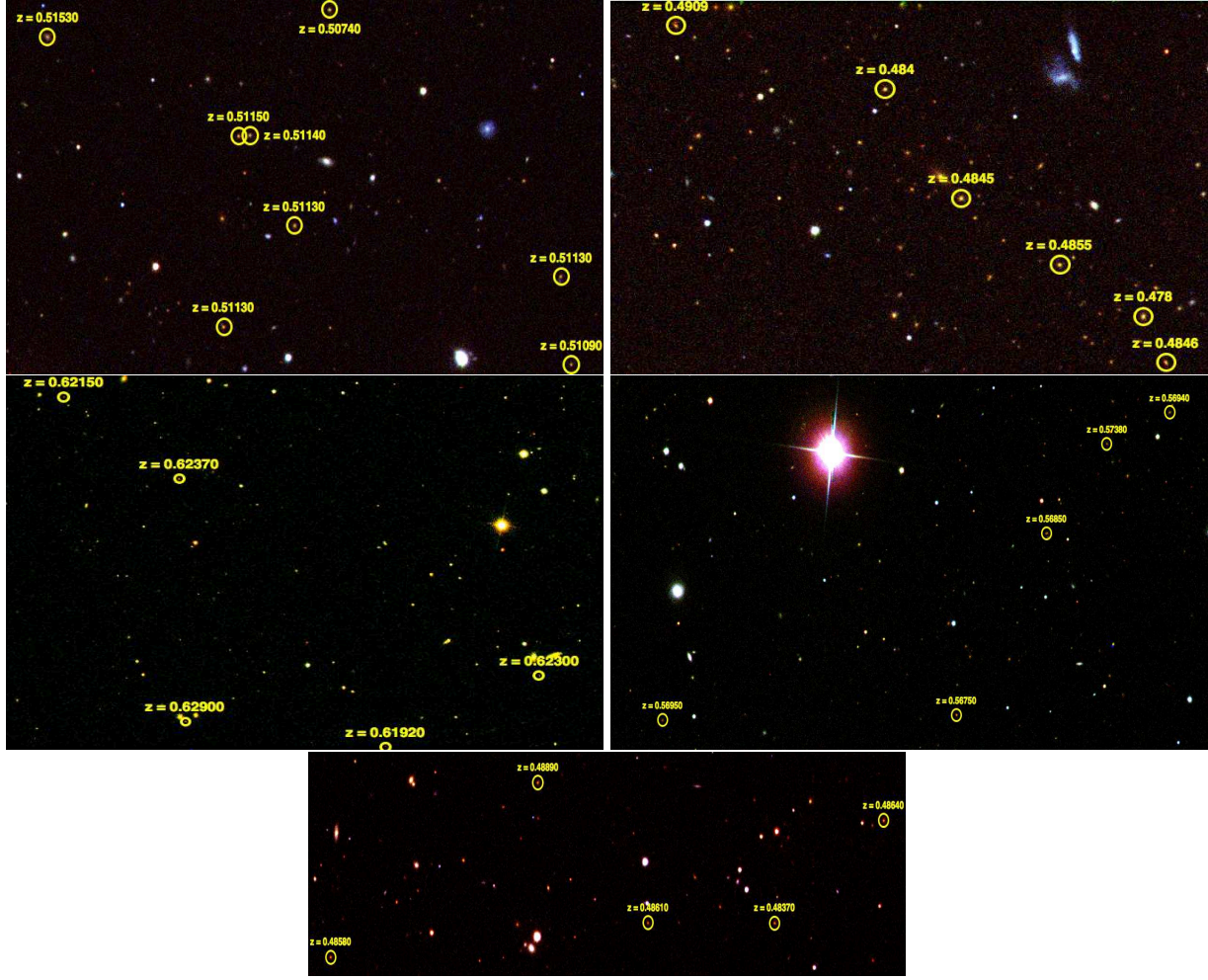
## 7 CONCLUSIONS

We have written an optical cluster finding algorithm based on that of Huchra & Geller (1982). Our dynamic friends-of-friends (DFoF) code uses a linking length that compensates for selection effects by changing size depending on the surface number density of galaxies at a given redshift.

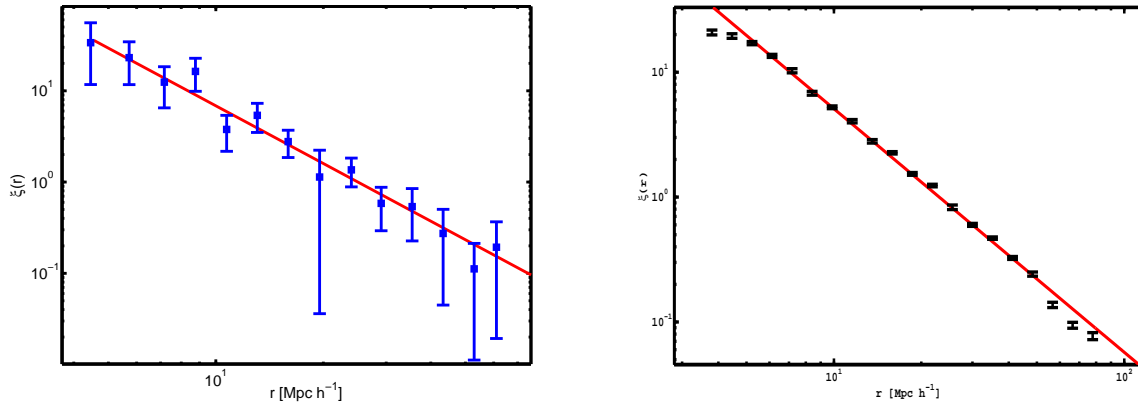
We produced a mock catalogue, which is representative of the 2SLAQ catalogue, using the Horizon  $4\pi$  simulation and the HOD prescription of Blake et al. (2008) in order to determine the linking parameters,  $R_{friend}(z)$  and  $v_{friend}(z)$ , in the DFoF code. The code was run with various combinations of the two parameters producing a set of distinct catalogues. We wrote a membership matching code in order to determine the completeness and purity of the resulting catalogues relative to the original mock haloes. Based on this analysis we chose values of  $R_{friend}(z = 0.5) = 0.87$  Mpc  $h^{-1}$  and  $v_{friend}(z = 0.5) = 900$  kms $^{-1}$ , which correspond to a catalogue that is 98% complete and 52% pure with the largest number of groups and clusters possible. Running the DFoF code with these values, we produced a catalogue of 313 groups and clusters containing 1152 member galaxies. The galaxy groups and clusters have an average velocity dispersion of  $\bar{\sigma}_v = 467.97$  kms $^{-1}$  and an average size of  $\bar{R}_{clt} = 0.78$  Mpc  $h^{-1}$ .

We tested the validity of our catalogue by obtaining SDSS galaxies in a 1deg<sup>2</sup> region around each cluster centre. We then subtracted the background signal from all the SDSS galaxies within 1 Mpc  $h^{-1}$  of the cluster centre and stacked the results in bins according to cluster redshift. Each of these

<sup>2</sup> This software is available from <http://healpix.jpl.nasa.gov/>



**Figure 13.** SDSS *g*, *r* and *i*-bands colour images of field around CL\_008 (top left panel), CL\_204 (top right panel), CL\_122 (middle left panel), CL\_024 (middle right panel) and CL\_038 (bottom panel). Yellow circles indicate the location of the cluster member galaxies, which are labeled with their individual redshifts.



**Figure 14.** Left panel: Two-point correlation function for the groups and clusters found in the 2SLAQ catalogue using the DFofF code (blue squares with error bars) and best-fit power law slope (red solid line). Right panel: Two-point correlation function for the mock halo catalogue (black squares with error bars) and best-fit power law slope (red solid line).  $\xi(r)$  was measured using the estimator of Landy & Szalay (1993).

bins was examined in SDSS photometric redshift space and colour-magnitude space. This analysis shows that our groups and clusters are reliable out to  $z \sim 0.6$ . The discrepancies beyond this range may be owing to the small number of groups and clusters detected at higher redshifts. Also, there may be some contamination effects in the histograms from the photometric redshift errors in the SDSS data.

We produce mass estimates for our catalogue using the cluster velocity dispersions according to equation 19. We tested the reliability of this approach by comparing mass estimates for clusters found in the 2SLAQ mock with the true masses of the mock haloes to which they are matched. We find that the mass estimates of our mock catalogue groups and clusters are for the most part a good fit to the true halo masses, however with large error bars. The deviations seen are the result of contaminating galaxies in the richness estimates and the small number of high mass clusters detected. In general we see that the range of cluster masses is a good match to known masses of massive clusters.

We analysed optical SDSS  $g$ ,  $r$  and  $i$ -band colour images of a selection of the our clusters, which span different redshifts. We observe that the galaxy members are distributed in a small region of space, on the sky and in redshift. We also see an overdensity of red galaxies around the cluster centres. This is strong evidence that the clusters are genuine. The distribution of LRGs in the clusters varies from spherical to elongated filamentary structures. This highlights an advantage of the percolation method in that it makes no prior assumptions about the cluster shapes, which allows us to detect some structures that other methods may not.

We test different clipping procedures on the 2SLAQ mock group and cluster sizes. This analysis indicated that the majority of the contamination in the cluster catalogue was attributed to groups with 3 or 4 members. We found that by clipping out clusters with  $R_{clt} > 0.11 \text{ Mpc } h^{-1}$  for  $N_{mem} = 3$  and  $R_{clt} > 0.49 \text{ Mpc } h^{-1}$  for  $N_{mem} = 4$  improves the purity from 52% to 88%, while reducing the completeness by only 4%, in the 2SLAQ mock catalogue. By applying this procedure to the real 2SLAQ group and cluster catalogue, we separated the clusters into ‘gold’ and ‘silver’ samples. Where the gold samples consists of clusters that passed the clipping procedure and are therefore the most reliable and the silver sample consists of clusters that failed the clipping procedure and may still be genuine, however less reliable. Out of the 313 total 2SLAQ groups and clusters, 70 are gold and the remaining 243 are silver.

Finally, we test the two-point correlation function of our cluster catalogue. We find a best-fitting power law model,  $\xi(r) = (r/r_0)^\gamma$ , with parameters  $r_0 = 24 \pm 4 \text{ Mpc } h^{-1}$  and  $\gamma = -2.1 \pm 0.2$ . The value of the reduced chi-squared is  $\chi^2_{red} = 0.94$ . These values are consistent with those of the mock halo catalogue and are in good agreement with those in literature (Nichol 2001).

Future surveys such as the Dark Energy Survey (DES), Euclid and Planck will images millions of galaxies across the whole sky. An abundance of photometric data will be obtained for each of these objects, however it will not be possible to obtain spectroscopic data for all of them. Therefore, it is important to develop reliable cluster finding techniques that utilise the photometric data that will be available. Photometric redshifts, for example, provide a useful way to probe the properties of galaxies along the line of

sight when spectroscopic data is not present. All of the results presented in the paper will be compared to those found in Farrens et al. (in prep), which will examine galaxy clusters in the 2SLAQ catalogue using photometric redshifts. This comparison will test the reliability of the DFOF code to detect structures using photometric data.

## ACKNOWLEDGEMENTS

The authors would like to thank Nic Ross for data and assistance with 2SLAQ correlation functions, Prof. Ofer Lahav for his advice and support, Ignacio Ferreras and Marcus Vinícius Costa Duarte for help regarding k-corrections, the DES clusters working group for discussions, John Deacon, Fabrizio Sidoli and Dugan Witherick for technical support, and Magda Vasta for helpful suggestions and tips. The authors also acknowledge the use of the UCL Legion High Performance Computing Facility, and associated support services, in the completion of this work. SF and CGS acknowledge support received from the Science and Technology Facilities Council and FBA acknowledges support received from the Royal Society via URF.

## REFERENCES

- Abell G. O., 1958, *ApJS*, 3, 211
- Bahcall N. A., Soneira R. M., 1983, *Apj*, 270, 20
- Blake C., Collister A., Lahav O., 2008, *MNRAS*, 385, 1257
- Borgani S., Plionis M., Kolokotronis V., 1999, *MNRAS*, 305, 866
- Botzler C. S., Snigula J., Bender R., Hopp U., 2004, *MNRAS*, 349, 425
- Bryan G. L., Norman M. L., 1998, *Apj*, 495, 80
- Cannon R., Drinkwater M., Edge A., Eisenstein D., Nichol R., Outram P., Pimbblet K., 2006, *MNRAS*, 372, 425
- Colberg J. M., White S. D. M., Yoshida N., MacFarland T. J., Jenkins A., Frenk C. S., Pearce F. R., Evrard A. E., Couchman H. M. P., Efstathiou G., Peacock J. A., Thomas P. A., The Virgo Consortium 2000, *MNRAS*, 319, 209
- Colless M., 1999, Royal Society of London Philosophical Transactions Series A, 357, 105
- Collins C. A., Guzzo L., Böhringer H., Schuecker P., Chincharini G., Cruddace R., De Grandi S., MacGillivray H. T., Neumann D. M., Schindler S., Shaver P., Voges W., 2000, *MNRAS*, 319, 939
- Couch W. J., Ellis R. S., MacLaren I., Malin D. F., 1991, *MNRAS*, 249, 606
- Dalton G. B., Maddox S. J., Sutherland W. J., Efstathiou G., 1997, Vizier Online Data Catalog, 728, 90263
- Davis M., Efstathiou G., Frenk C. S., White S. D. M., 1985, *Apj*, 292, 371
- Driver S. P., Baldry I. K., Bamford S., Bland-Hawthorn J., Bridges T., Cameron E., Conselice C., J. C. W., Croom S., Cross N. J. G., Driver S. P., Dunne L., Eales S., Edmondson E., Ellis S. C., Frenk C. S., Graham A. W., Jones H., 2008, ArXiv e-prints
- Efstathiou G., Frenk C. S., White S. D. M., Davis M., 1988, *MNRAS*, 235, 715



Eisenstein D. J., Hut P., 1998, *Apj*, 498, 137  
Gal R. R., 2006, ArXiv Astrophysics e-prints  
Geller M. J., Huchra J. P., 1983, *ApJS*, 52, 61  
Gladders M. D., Yee H. K. C., 2000, *Aj*, 120, 2148  
Goto T., Sekiguchi M., Nichol R. C., Bahcall N. A., Kim R. S. J., Annis J., Ivezić Ž., Brinkmann J., Hennessy G. S., Szokoly G. P., Tucker D. L., 2002, *Aj*, 123, 1807  
Haiman Z., Mohr J. J., Holder G. P., 2001, *Apj*, 553, 545  
Huchra J., Davis M., Latham D., Tonry J., 1983, *ApJS*, 52, 89  
Huchra J. P., Geller M. J., 1982, *Apj*, 257, 423  
Jones D. H., Peterson B. A., Colless M., Saunders W., 2006, *MNRAS*, 369, 25  
Kim R. S. J., Kepner J. V., Postman M., Strauss M. A., Bahcall N. A., Gunn J. E., Lupton R. H., Annis J., Nichol R. C., Castander F. J., Brinkmann J., Brunner R. J., Connolly A., Csabai I., Hindsley R. B., Ivezić Ž., Vogeley M. S., York D. G., 2002, *Aj*, 123, 20  
Koester B. P., McKay T. A., Annis J., Wechsler R. H., Evrard A. E., Rozo E., Bleem L., Sheldon E. S., Johnston D., 2007, *Apj*, 660, 221  
Kravtsov A. V., Berlind A. A., Wechsler R. H., Klypin A. A., Gottlöber S., Allgood B., Primack J. R., 2004, *Apj*, 609, 35  
Landy S. D., Szalay A. S., 1993, *Apj*, 412, 64  
Lidman C. E., Peterson B. A., 1996, *Aj*, 112, 2454  
Lopes P. A. A., de Carvalho R. R., Gal R. R., Djorgovski S. G., Odewahn S. C., Mahabal A. A., Brunner R. J., 2004, *Aj*, 128, 1017  
Lupton R., 1993, Statistics in theory and practice. Princeton, N.J.: Princeton University Press, —c1993  
Maddox S. J., Efstathiou G., Sutherland W. J., Loveday J., 1990, *MNRAS*, 243, 692  
Miller C. J., Batuski D. J., Slinglend K. A., Hill J. M., 1999, *Apj*, 523, 492  
Navarro J. F., Frenk C. S., White S. D. M., 1996, *Apj*, 462, 563  
Nichol R. C., 2001, ArXiv Astrophysics e-prints  
Peebles P. J. E., 1993, Physics Today, 46, 87  
Postman M., Lubin L. M., Gunn J. E., Oke J. B., Hoessel J. G., Schneider D. P., Christensen J. A., 1996, *Aj*, 111, 615  
Press W. H., Schechter P., 1974, *Apj*, 187, 425  
Prunet S., Pichon C., Aubert D., Pogossyan D., Teyssier R., Gottloeber S., 2008, *ApJS*, 178, 179  
Ramella M., Boschini W., Fadda D., Nonino M., 2001, *A&A*, 368, 776  
Ramella M., Geller M. J., Pisani A., da Costa L. N., 2002, *Aj*, 123, 2976  
Ross N. P., da Ângela J., Shanks T., Wake D. A., Cannon R. D., 2007, *MNRAS*, 381, 573  
Sawangwit U., Shanks T., Abdalla F. B., Cannon R. D., Croom S. M., Edge A. C., Ross N. P., Wake D. A., 2009, ArXiv e-prints  
Shapley H., Ames A., 1932, Annals of Harvard College Observatory, 88, 41  
Sheth R. K., Diaferio A., 2001, *MNRAS*, 322, 901  
Teyssier R., 2002, *A&A*, 385, 337  
Teyssier R., Pires S., Prunet S., Aubert D., Pichon C., Amara A., Benabed K., Colombi S., Refregier A., Starck J., 2009, *A&A*, 497, 335  
Tinker J. L., 2007, *MNRAS*, 374, 477

van Haarlem M. P., Frenk C. S., White S. D. M., 1997, *MNRAS*, 287, 817  
Weller J., Battye R. A., 2003, New Astronomy Review, 47, 775  
Weller J., Battye R. A., Kneissl R., 2002, Physical Review Letters, 88, 231301  
York D. G., Adelman J., Anderson Jr. J. E., Anderson S. F., Annis J., Bahcall N. A., 2000, *Aj*, 120, 1579

## APPENDIX A: CLUSTER CATALOGUE

The catalogue contains 313 clusters with the following properties: identifier, number of galaxy members (or richness), right ascension, declination, redshift, velocity dispersion (in  $\text{kms}^{-1}$ ), radial size (in  $\text{Mpc } h^{-1}$ ), estimated mass (in  $\log_{10} M_{\odot}$ ) and sample to which it belongs (G: gold ; S: silver). A sample of the first 10 clusters is shown in table-A1.



**Table A1.** Cluster Catalogue Sample

$N_{\bar{o}}$	ID	$N_{mem}$	RA	Dec	$z$	$\sigma_v$ (kms $^{-1}$ )	$R_{clt}$ (Mpc $h^{-1}$ )	$\log_{10} M_{est}$ ( $M_{\odot}$ )	Sample
1	CL_0008	011	152.85693	-0.00323	0.51165	0562.51	02.08	14.3051	G
2	CL_0031	010	210.59379	-0.55691	0.44007	1178.60	02.00	15.2916	G
3	CL_0060	009	035.42705	-0.22677	0.59094	0784.58	01.74	14.7136	G
4	CL_0140	008	342.39005	-0.44521	0.49785	0846.21	01.27	14.8415	G
5	CL_0204	007	158.86053	-0.02893	0.48510	1108.13	00.65	15.1969	G
6	CL_0038	007	191.56150	-0.51722	0.48679	0511.13	01.66	14.1882	G
7	CL_0051	007	190.76950	+0.03930	0.46779	0786.90	01.26	14.7564	G
8	CL_0122	007	161.24470	-0.30396	0.62371	0851.02	01.85	14.8094	G
9	CL_0064	006	202.19576	-0.35692	0.52837	0381.83	00.73	13.7950	G
10	CL_0024	006	343.37305	-0.49057	0.56982	0591.89	01.52	14.3531	G